DTIC FILE COPY

④

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AI Memo 1007	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Qualitative Depth and Shape from Stereo, in Agreement with Psychophysical Evidence		5. TYPE OF REPORT & PERIOD COVERED memorandum
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Daphna Weinshall		8. CONTRACT OR GRANT NUMBER(s) N00014-85-K-0124
9. PERFORMING ORGANIZATION NAME AND ADDRESS Artificial Intelligence Laboratory 545 Technology Square Cambridge, MA 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE December 1987
		13. NUMBER OF PAGES 20
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report)
		16a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Distribution is unlimited.

DTIC ELECTE JUN 2 1 1988 S D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

None

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

stereo vision,
qualitative depth,
induced effect,
motion,

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

ABSTRACT: This paper concentrates on the problem of obtaining depth information from binocular disparities. It is motivated by the fact that implementing registration algorithms and using the results for depth computations is hard in practice with real images due to noise and quantization errors. We will show that qualitative depth information can be obtained from stereo disparities with almost no computations, and with no prior knowledge (or computation) of camera parameters. The only constraint is that the epipolar plane of

the fixation point includes the X—axes of both cameras. We derive two expressions that order all matched points in the images in two distinct depth-consistent fashions from image coordinates only. One is a tilt-related order λ, which depends only on the polar angles of the matched points, the other is a depth-related order χ. Using λ for tilt estimation and point separation (in depth) demonstrates some anomalies and unusual characteristics that have been observed in psychophysical experiments, most notably the "induced size effect". Furthermore, the same approach can be applied to estimate some qualitative behavior of the normal to the surface of any object in the field of view. More specifically, one can follow changes in the curvature of a contour on the surface of an object, with either x- or y-coordinate fixed.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY
and
CENTER FOR BIOLOGICAL INFORMATION PROCESSING
WHITAKER COLLEGE

Qualitative Depth and Shape from Stereo, in Agreement with Psychophysical Evidence

Daphna Weinshall

ABSTRACT: This paper concentrates on the problem of obtaining depth information from binocular disparities. It is motivated by the fact that implementing registration algorithms and using the results for depth computations is hard in practice with real images due to noise and quantization errors. We will show that qualitative depth information can be obtained from stereo disparities with almost no computations, and with no prior knowledge (or computation) of camera parameters. The only constraint is that the epipolar plane of the fixation point includes the X−axes of both cameras. We derive two expressions that order all matched points in the images in two distinct depth-consistent fashions from image coordinates only. One is a tilt-related order λ, which depends only on the polar angles of the matched points, the other is a depth-related order χ. Using λ for tilt estimation and point separation (in depth) demonstrates some anomalies and unusual characteristics that have been observed in psychophysical experiments, most notably the "induced size effect". Furthermore, the same approach can be applied to estimate some qualitative behavior of the normal to the surface of any object in the field of view. More specifically, one can follow changes in the curvature of a contour on the surface of an object, with either x- or y-coordinate fixed.

Introduction

Research in early vision regarding stereo seems to be concerned mainly with the correspondence problem, namely, finding the right matching between points on the left and right images. Obtaining exact depth values from a stereo pair has been considered a simple exercise, whose solution is well known, though might involve some tedious but straightforward computations. Thus, it has been implicitly assumed that the final goal of stereo algorithms is to compute an exact depth map using disparity values. The following observations suggest, however, that the depth computation from disparity values is not necessarily straightforward or even feasible, and that more qualitative depth information may be easier to obtain and more robust.

First, the depth computation problem reduces to a simple trigonometric formula when the parameters of the cameras, or the eyes, are known. When they are not known, a scheme to compute the camera's parameters from a number of conjugate points (that is, matched pairs of points from the different images) has been devised, involving the solution of a set of nonlinear equations (see for instance Horn, 1986). Since the problem has no closed-form solution, and since the data are not precise, a solution is found using iterative methods that minimize the sum of the squares of the errors. In practice, however, this approach is very difficult to implement, since the parameters of the cameras must be obtained from data with error on the order of magnitude of the disparity values, which are the raw material used for depth computation (e.g., error due to pixel quantization). In other words, the registration problem (namely, finding parameters for the camera's calibration) is much more difficult than just computing depth from disparity values. Less general methods to perform camera calibration have also been devised, see Prazdny (1981) and Longuet-Higgins (1981).

The other observation originates from biological vision. It seems that human vision does not necessarily obtain exact depth values from stereo disparity information alone, see, e.g., Foley (1977) and Foley & Richards (1972). Rather, stereo disparity seems to be used mainly in obtaining qualitative depth information about objects in the field of view. Estimation of the magnitude of this relative depth is possibly dependent on extraretinal estimation of some physical parameters like the angle of convergence of the eyes. For example, whether looking at stereograms with crossed or uncrossed eyes affects only the extraretinal perception of the angle of convergence of the eyes, not the disparity values. It also results in a different perception of the depth of the central square in a simple random-dot stereogram (where a central square in one image has a constant shift with respect to the other).

The purpose of this paper is to exploit the geometry of the situation, where a scene is viewed from two different angles, to obtain insight into the above problem. It will be shown that qualitative relative depth information (order) of various kinds can be obtained from conjugate points alone easily and reliably, involving almost no computations and independent of the camera's parameters. These orders will demonstrate some anomalies that are observed in human psychophysics and presently lack other straightforward explanations.
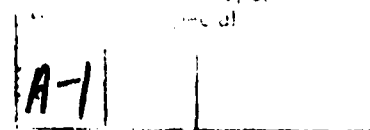
One anomaly is the induced size effect, where a distortion of one image by stretching the Y-axis (the vertical axis) of that image produces a tilt-impression similar to that produced by stretching the X-axis (the horizontal axis) of the other image by the same amount. Whereas the tilt impression caused by stretching the X-axis has a simple geometrical explanation, the reversed tilt impression caused by stretching the Y-axis has no geometrical logic behind it, and has therefore been called an induced effect. Induced, since it is as if the unrealistically magnified Y-axis induces the shrinking of the X- and Y-axis as a compensation, which has the similar geometrical interpretation as stretching the X-axis of the other image. This effect, first reported by Ogle (1938), has stimulated extensive research, see Arditi et al. (1981), Mayhew (1982), Longuet-Higgins (1982), Mayhew & Longuet-Higgins (1982), and Rogers & Koenderink (1986).

It will also be shown that some qualitative shape information can be obtained from image coordinates only. More specifically, one can follow changes in the curvature of a contour on the surface of any object in the field of view, with either x- or y-coordinate fixed.

The exact order expressions will scale in proportion to the angle of convergence between the two cameras. The exact relative depth can be computed from these orders using few matched points and some approximated numeric scheme, or using more than two images. Alternatively, it can be estimated from some external estimation of the physical quantities involved, namely the angle of convergence and the angle of gaze, in agreement with the psychophysical theory suggested in Foley (1977).

Basic Geometry

Given two cameras, assume that the optical axes intersect at the fixation point. Also, assume that the epipolar plane of the fixation point (the plane through the optical axes of both cameras and their baseline, henceforth "base plane") includes the X-axes of both cameras (which are, therefore, epipolar lines by definition). Let us define the following coordinate system (see figure 1): let the fixation point be the origin, the base plane (which passes through this point) be the $X - Z$ plane, and the line perpendicular to this plane through the origin be the Y-axis. On the $X - Z$ plane, the optical axes of both cameras intersect in the origin and create an angle 2μ between them. Let the Z-axis be the angle-bisector of 2μ, and the X-axis perpendicular to the Z-axis in the $X - Z$ plane. This system is very similar to the cyclopean coordinate system commonly used in the literature, with the exceptions that the angle-bisector is replaced by the median to the interocular line and the origin is translated to the mid point of the interocular line. A similar system can be defined for the case of motion, if the fixation point is kept constant. That is, the observer follows the same point with his eyes. This is more typical of human vision than machine vision.

For a given point $P = (x, y, z)$, let α denote the angle of tilt - the polar angle of its projection on the $X - Z$ plane ($\alpha = \arctan(\frac{x}{z})$). Let β denote the angle of slant - the
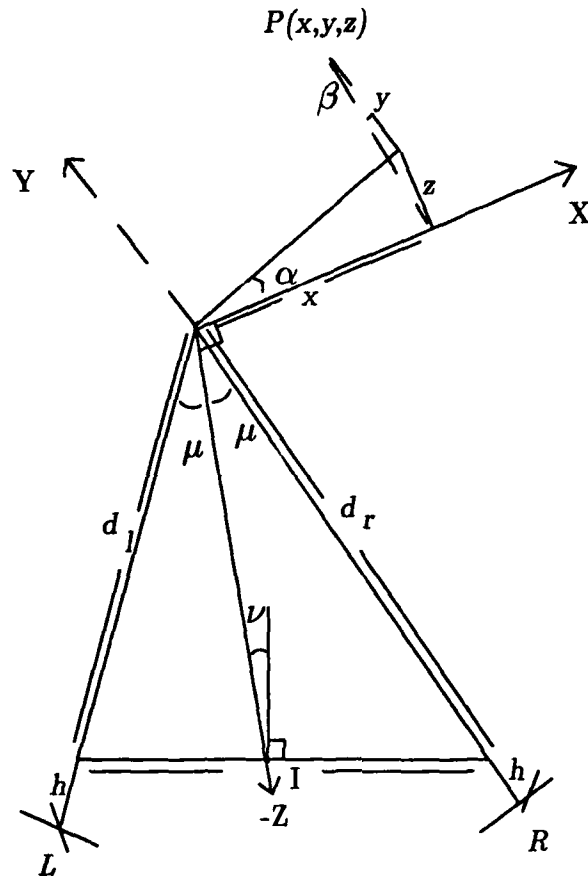
Figure 1. The base plane $(X - Z)$ viewed from above, with two cameras.

polar angle of its projection on the Y-Z plane ($\beta = \arctan(\frac{z}{y})$). Thus P can be also written as $P = (\frac{z}{\tan \alpha}, \frac{z}{\tan \beta}, z)$, where z is its depth relative to the fixation point in the above coordinate system. Let (x_l, y_l) and (x_r, y_r) be the Cartesian coordinates of the projection of P on the left and right images respectively. Using polar coordinates, the two projections can be written as (r_l, ϑ_l) and (r_r, ϑ_r) respectively. Let $\lambda = \frac{\cot \vartheta_r}{\cot \vartheta_l}$. Then the following can be shown to hold (see appendix):

$$\tan \alpha = \frac{1}{\tan \mu} \cdot \frac{\lambda - 1}{\lambda + 1} \qquad , \tag{1}$$

$$\tan \beta = \frac{\cot \vartheta_r - \cot \vartheta_l}{2 \sin \mu} \tag{2}$$

Thus, the two angles α and β depend only on the angle of convergence and the polar

3

angles of the conjugate points. It can be shown that the polar angles are preserved under projection, through any point on the optical axis, onto either a spherical body (like the eye) or a planar one (a camera). There is no dependence on other parameters of the cameras (which could be different ones), their relative positions, or the angle of gaze. Equation (1) will be used in the next section to obtain an order on all matched image points in each visual hemifield according to their tilt. This order is independent of the camera parameters and demonstrates psychophysical anomalies like the induced effect and others. Equation (2) will be used to obtain an expression for the relative depth z. However, this expression will depend on the camera's parameters like focal length and interocular distance, and the angle of gaze. A parameter-independent relative depth order can be obtained from this expression. It will hold for small angles of convergence 2μ, as will be discussed in a succeeding section.

Tilt-related Order

From (1) it immediately follows that α is a monotonic increasing function of λ for a fixed configuration of the cameras. Thus, the ratio

$$\frac{\lambda_i}{\lambda_j} = \frac{\cot \vartheta_r^i}{\cot \vartheta_l^i} \bigg/ \frac{\cot \vartheta_r^j}{\cot \vartheta_l^j}$$

gives qualitative relative distance information on any two points i and j in each visual hemifield in the following sense: if the ratio is greater than 1, meaning $\alpha_i > \alpha_j$, then a separating plane between points i and j through the fixation point and perpendicular to the base plane will leave point j and the viewer on one of its sides, and point i on the other side "further away" from the viewer. In other words, λ defines an order on all the matched points in a given hemifield. This order corresponds to Euclidean distance if points i and j are approximately on the same line of sight from the viewer, namely about the same image x-coordinate.

Note that if λ is constant on all points segmented as belonging to the central object, then it is a planar object with some tilt towards or away from the viewer, according to the sign of $\lambda - 1$. Moreover, since

$$\vartheta_l \gtreqless \vartheta_r \iff \frac{\cot \vartheta_r}{\cot \vartheta_l} \gtreqless 1 \iff \alpha \gtreqless 0^o,$$

it follows that $\vartheta_l - \vartheta_r$ also gives a qualitative estimation to the tilt of a point P relative to the fixation point. If the fixation point is at the same distance from both cameras, this estimate would indicate whether P, relative to the fixation point, is tilted away from the camera's baseline ($\alpha > 0^o$), "parallel" to the baseline ($\alpha \approx 0^o$), or towards the camera's baseline ($\alpha < 0^o$).

The order expression λ has been defined as a function of the polar angle ϑ only in both eyes. This is especially convenient since the polar angle is preserved under projection onto either a spherical body (the eye) or a planar body (a camera). However, it might prove

useful to examine λ as a function of the Cartesian coordinates (x_l, y_l) and (x_r, y_r) in both images, assuming planar projection. In this case:

$$
\begin{aligned}
\ln \lambda &= \ln \frac{\cot \vartheta_r}{\cot \vartheta_l} = \ln \frac{x_r/x_l}{y_r/y_l} \\
&= (\ln x_r - \ln x_l) - (\ln y_r - \ln y_l) \\
&= \Delta(\ln x) - \Delta(\ln y).
\end{aligned}
\tag{4}
$$

In other words, if any matching algorithm is applied to the output images of the transformation $T : (x, y) \rightarrow (\ln x, \ln y)$ performed on the original images, and the disparity vector (Δ_x, Δ_y) is then computed in the usual way, then the difference $\Delta_x - \Delta_y = \ln \lambda$ is an order of the same type as λ, with no need for any additional computation. The transformation T is indeed singular near the vertical and horizontal meridians. However, there is evidence from neurophysiology for the existence of a related transformation between the retina and $V1$, the complex log, which is singular at the origin (see Schwartz, 1984).

One prediction of the order λ is the "induced effect", the psychophysical effect where a distortion of one image by stretching the Y-axis (the vertical axis) of that image produces a tilt impression similar to that produced by stretching the X-axis (the horizontal axis) of the other image by the same amount (see introduction). Evidently, the decision rule specified above exhibits the same qualitative behavior since it involves only terms of the form $\frac{y}{x}$. Hence multiplying the Y-axis by some number has the same effect as multiplying the X-axis by its inverse. Thus an induced effect is a natural side-effect of using the relation $\frac{\cot \vartheta_r}{\cot \vartheta_l}$, which does not depend on any assumptions and approximations, or the complete recovery of all depth-related parameters of the scene. On the other hand, some other computational explanations to this effect, e.g. Mayhew & Longuet-Higgins (1982) and Mayhew (1982), obtain the induced effect as a by-product of a specific approximation scheme and a tedious numerical computation; it does not result from an exact solution of the disparity equations. Another computational explanation (Arditi et al., 1981) suggests that a distortion occurs in the matching stage, assuming matching is done along horizontal lines only.

Motion also shows an illusion similar to the induced effect (see Rogers & Koenderink, 1986). In this case, observers reported that a fronto-parallel surface appeared to be tilting in depth with the right-hand side apparently closer than the left when the monocular image was progressively magnified with head movement to the right and vice versa. λ can account for this phenomenon. Moreover, in this case there is an additional effect - a perceived forward/backward motion. This could possibly be accounted for by the angle of gaze ν. As will be shown later,

$$
\tan \nu = \frac{1}{\tan \mu} \cdot \frac{1 - \frac{y_r(x=0)}{y_l(x=0)}}{1 + \frac{y_r(x=0)}{y_l(x=0)}}.
$$

Thus, a distortion of the y-axis in one image will distort ν (a distortion of the x-axis will not affect ν, however). The angle of gaze ν can be used to obtain the direction of motion in the following way (see figure 2): ν is 0 (the true angle of gaze) for motion from point

1 to point 2. Positive ν, the computed angle of gaze, implies motion from 1 to 2′, that is, backward movement of the head in addition to its left to right movement. Since the head does not move, the object is perceived as moving backwards.
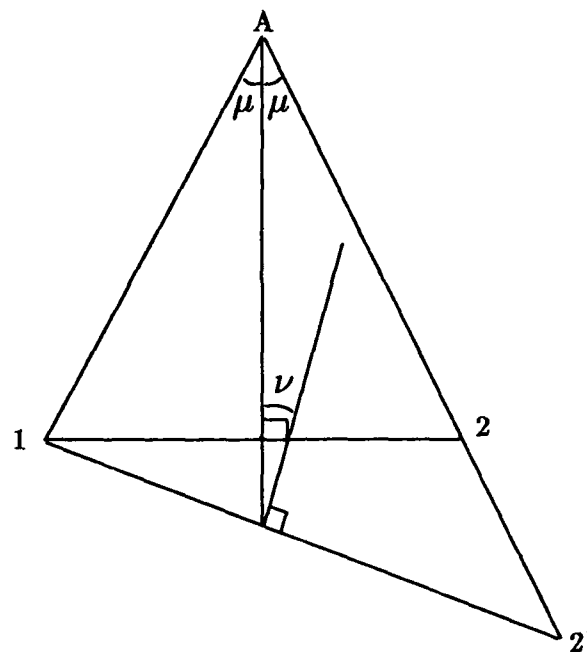

Figure 2 False positive angle of gaze induces perception of backward motion (point 2′), whereas the true value 0 shows no such motion (point 2), see text.

The quantitative aspect of this order, namely, obtaining numerical values for α and μ from λ, will be discussed in a succeeding section. However, since computing λ involves computing ratios of the image's x-coordinates and y-coordinates, one should expect problems near the meridians. The horizontal meridian is of special interest since the matching algorithm should give relatively good results on this meridian. In this respect, it is interesting to note that human observers also demonstrate deteriorating performance near the meridians, especially near the horizontal meridian (Arditi, 1982). This deterioration is demonstrated by a smaller probability for a correct detection of the tilt of an oblique line when either the x- or y-axis is magnified, and when the angle between the oblique line and the horizontal meridian is around either $0°$ or $90°$. Compensation for errors near the horizontal meridian is further discussed in the section on error analysis.

Depth-related Order

From equation (2) one can obtain an explicit expression for the depth z of a point relative to the fixation point (the origin). First, note that (2) implies

$$z = (\cot \vartheta_r - \cot \vartheta_l) \cdot \frac{y}{2 \sin \mu}. \tag{5}$$

Thus, $\chi^y = (\cot \vartheta_r - \cot \vartheta_l)$ gives exact depth order on all the points in space with some constant height y over the base plane. It follows that this order is most useful to compare points that differ mainly in their x-coordinate with y approximately the same. The previous order λ, on the other hand, was most useful to compare points that differ mainly in their y-coordinate with x approximately the same. This fact stands in agreement with much psychophysical evidence on the anisotropy between the vertical and horizontal dimensions.

Next, let us derive an expression for z that depends only on scene and camera parameters. In the appendix it is shown that, for ν the angle of gaze, I the interocular distance, and h the focal length of the cameras,

$$y = \left\{ \frac{I \cos(\mu - \nu)}{\sin 2\mu} - x \sin \mu + z \cos \mu \right\} \cdot \frac{y_r}{h}, \tag{6}$$

or

$$y = \left\{ \frac{I \cos(\mu + \nu)}{\sin 2\mu} + x \sin \mu + z \cos \mu \right\} \cdot \frac{y_l}{h}. \tag{6'}$$

Substituting (6) in (5) for a point in the right hemifield, ((6') will be used otherwise), gives:

$$z = \frac{I \cos(\mu - \nu)/\sin 2\mu}{\left\{ \frac{\sin \mu}{x_r - \frac{y_r}{y_l} x_l} \left[2h + \tan \mu (x_r + \frac{y_r}{y_l} x_l) \right] - \cos \mu \right\}}.$$

Thus, for an angle of convergence 2μ small enough so that $2h \gg |\tan \mu (x_r + \frac{y_r}{y_l} x_l)|$, we obtain a relative depth order χ on all the points in the visual field, where

$$\chi = x_r - \frac{y_r}{y_l} x_l.$$

As will be shown in the section on error analysis, $\frac{y_r}{y_l} = 1 + O(\mu)$. Likewise, since the field of view is mechanically bounded by some $2\xi < 180^o$, it follows that $x < h \tan \xi$. Thus, a sufficient condition for the appropriateness of χ, to a first order in μ, is $1 \gg \tan \mu \cdot \tan \xi$. If $2\xi \leq 90^o$, which is a reasonable upper bound, then it is sufficient if $1 \gg \tan \mu$, or $2\mu \ll 90^o$. To illustrate, the distance to the point of fixation should be much greater than 3 cm for an average person looking straight ahead, possibly 30 cm or more.

We have obtained, then, a relative depth order that is the traditional x-disparity corrected for non zero vergence (angle of gaze ν not 0) and some field location (x-coordinate) distortion. However, for a fixed convergence angle 2μ, this order has some distortion relative to the physical relative depth, which increases with the horizontal distance from the point of fixation (the x-coordinate).

Qualitative Shape from Stereo

The triple (α, β, z) as a point representation, and equations (1) and (2), turn out to be useful for surface normal analysis. For any two points P_1 and P_2, where $\vec{P}_1 = z_1(\frac{1}{\tan \alpha_1}, \frac{1}{\tan \beta_1}, 1)$ and $\vec{P}_2 = z_2(\frac{1}{\tan \alpha_2}, \frac{1}{\tan \beta_2}, 1)$, let $\vec{N} = \vec{P}_1 \times \vec{P}_2$. $\vec{N}$ is perpendicular to $(\vec{P}_1 - \vec{P}_2)$. (It is actually proportional to the normal to the plane passing through P_1, P_2, and the fixation point.) After some calculations, it can be shown that

$$\vec{N} = z\left(\frac{\cot \beta_1 - \cot \beta_2}{\cot \alpha_1 \cot \beta_2 - \cot \alpha_2 \cot \beta_1} , \frac{\cot \alpha_2 - \cot \alpha_1}{\cot \alpha_1 \cot \beta_2 - \cot \alpha_2 \cot \beta_1} , 1\right)$$

$$= z\left(\frac{1}{\tan \mu} f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2) , \frac{1}{\sin \mu} g(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2) , 1\right),$$

where

$$f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2) = \frac{(\cot \vartheta_l^2 - \cot \vartheta_l^1) - (\cot \vartheta_r^2 - \cot \vartheta_r^1)}{(\cot \vartheta_r^2 - \cot \vartheta_r^1) + (\cot \vartheta_l^2 - \cot \vartheta_l^1)}$$

$$g(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2) = \frac{\cot \vartheta_r^2 \cot \vartheta_l^1 - \cot \vartheta_l^2 \cot \vartheta_r^1}{(\cot \vartheta_r^2 - \cot \vartheta_r^1) + (\cot \vartheta_l^2 - \cot \vartheta_l^1)}.$$

Thus, as long as $f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ and $g(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ remain constant, which can be determined from image coordinates only, the points are coplanar (among themselves and with the fixation point), or the object at the center of gaze is planar. Note that λ is obtained from f when $\cot \vartheta_l^2 = \cot \vartheta_r^2 = 0$ ($g = 0$ then).

Moreover, for any object it is possible to obtain qualitative information about its surface along any contour, with either x or y fixed. Take a contour on the surface with some fixed y-coordinate, and let P_1 and P_2 be two points on it. Since the y-coordinate of $\vec{P}_1 - \vec{P}_2$ is 0, the projection of $\vec{N}$ on the $X - Z$ plane, $\vec{n} = z(\frac{1}{\tan \mu} f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2) , 1)$, is perpendicular to the projection of $\vec{P}_1 - \vec{P}_2$. Thus, for fixed y, the one dimensional boundary contour is convex when $f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ increases with increasing x, concave when $f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ decreases, and linear when $f(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ remains constant. Note that χ^y can be obtained from f since the sign of f determines relative depth between two points with fixed y-coordinate. The same qualitative description can be obtained for any boundary contour with fixed x from following $g(\vartheta_l^1, \vartheta_r^1, \vartheta_l^2, \vartheta_r^2)$ with increasing y, x fixed. This qualitative description depends on image coordinates only, more specifically on polar angles of the conjugate points. Thus it predicts an "induced effect" for planes that do not include the point of fixation. Obtaining this description is not trivial, though, since such a contour in the world coordinate system will be usually mapped to an oblique line in the image plane due to convergence.

In the general case, the normal to a plane passing through three points in space P_1, P_2, and P_3 depends on the image coordinates and the angle of convergence μ in a more complex way, so that μ should be known to compute the (exact) 3-D normal. (The focal length h should be known as well.) However, if $\vec{P}_1 - \vec{P}_2 = (0, y, z)$ and $\vec{P}_2 - \vec{P}_3 = (x, 0, z)$ or vice versa, the normal can be computed from the above argument (up to a scaling factor of the x- and y-coordinates, depending on μ).

8

Alternatively, one can estimate the normal to the plane passing between three points P_1, P_2, and P_3 to a first order in μ and the x-disparity

$$\frac{x_r - \frac{y_r}{y_l}x_l}{2h\sin\mu}.$$

In this case, after substituting $K_{\mu\nu h} \cdot (x_r - \frac{y_r}{y_l}x_l)$ as an approximation for z, where $K_{\mu\nu h}$ is some constant that depends on μ, ν, and h, one gets an expression for the general normal $\vec{N}_G$:

$$
\begin{aligned}
\vec{N}_G =& (\vec{P}_1 - \vec{P}_2) \times (\vec{P}_2 - \vec{P}_3) \\
\approx& W_{\mu\nu h}(V_{\mu\nu h} F(x_r^1, x_l^1, y_r^1, y_l^1, x_r^2, x_l^2, y_r^2, y_l^2, x_r^3, x_l^3, y_r^3, y_l^3), \\
& U_{\mu\nu h} G(x_r^1, x_l^1, y_r^1, y_l^1, x_r^2, x_l^2, y_r^2, y_l^2, x_r^3, x_l^3, y_r^3, y_l^3), \ 1),
\end{aligned}
$$

where $W_{\mu\nu h}$, $V_{\mu\nu h}$, and $U_{\mu\nu h}$ are some constants which depend on μ, ν and h. $F()$ and $G()$ are some functions of images coordinates only. Once again, one can verify planarity of surfaces of objects in the field of view when F and G remain constant.

Numerical Computation

Let us compute the exact tilt and depth value to a first order in the convergence angle 2μ, following Mayhew & Longuet-Higgins' (1982) method to compute tilt and slant of a plane through the fixation point. The following scheme, however, will be simpler and involve less and more rigorous assumptions (we shall only assume small 2μ as implied above). Since, to a first order in μ, $\tan 2\mu \approx \frac{I\cos(\nu)}{R}$, where R is the distance between the fixation point and the midpoint of the interocular line (the nose), our computations will be to a first order in $(\frac{1}{R})$.

Let (x, y) and (x', y') denote the image coordinates of a certain point in space on the two cameras, respectively. Let $\hat{\alpha}$ and $\hat{\beta}$ denote the parameters of a plane that passes through a given point in space and the fixation point in the above coordinate system, so that $Z = \hat{\alpha}X + \hat{\beta}Y$. Thus $\hat{\alpha}$ is $\tan(\alpha)$ in the previous notations if $\hat{\beta} = 0$ and $\hat{\beta}$ is $\tan(\beta)$ if $\hat{\alpha} = 0$.

Then, to a first order in μ, we have (Longuet-Higgins & Prazdny, 1980)

$$
\begin{aligned}
\Delta x = x' - x =& [(\hat{\alpha}\cos(\nu) + \sin(\nu))x + \hat{\beta}\cos(\nu)y \\
& + (\cos(\nu) - \hat{\alpha}\sin(\nu))x^2 - \hat{\beta}\sin(\nu)xy] \cdot I/R, \\
\Delta y = y' - y =& [\sin(\nu)y + (\cos(\nu) - \hat{\alpha}\sin(\nu))xy \\
& - \hat{\beta}\sin(\nu)y^2] \cdot I/R.
\end{aligned}
\tag{7}
$$

(The coordinate system used to obtain (7) is the cyclopean coordinate system. This, however, does not change the results when changing to our coordinate system since the angle-bisector and the median are the same line to a first order in μ and the translation of the origin has been taken into account in the definition of the target plane.)

9

For a given point in space, one can, in the more interesting cases, take the plane passing through it and the fixation point which is perpendicular to the base plane, for which $\hat{\beta} = 0$. This plane would be determined only by $\hat{\alpha}$ (the plane perpendicular to the base plane, which is usually unique). Thus, we have

$$\frac{\Delta y}{y} = [\sin(\nu) + (\cos(\nu) - \hat{\alpha}\sin(\nu))x] \cdot I/R$$

$$= [\tan(\nu) + (1 - \hat{\alpha}\tan(\nu))x] \cdot \tan(2\mu). \tag{8}$$

Let (x_1 , y_1 , Δx_1 , Δy_1) be the coordinates of a point on the vertical meridian, so that $x_1 \approx 0$. Then we have

$$\frac{\Delta y_1}{y_1} = \tan(\nu) \cdot \tan(2\mu).$$

(Recall that $\frac{1 - \frac{y_1'}{y_1}}{1 + \frac{y_1'}{y_1}} = \tan(\nu) \cdot \tan(\mu)$ always.)

Let (x_2 , y_2 , Δx_2 , Δy_2) be the coordinates of a point with $\hat{\alpha} \approx 0$. Such a point, if it exists, can be easily identified since it satisfies $\frac{y'}{y} \approx \frac{x'}{x}$. Then we have:

$$x_2 \cdot \tan(2\mu) = \frac{\Delta y_2}{y_2} - \tan(\nu) \cdot \tan(2\mu) = \frac{\Delta y_2}{y_2} - \frac{\Delta y_1}{y_1} = \frac{y_2'}{y_2} - \frac{y_1'}{y_1}.$$

In other words,

$$\tan(2\mu) = \frac{1}{x_2} \cdot \left[\frac{y_2'}{y_2} - \frac{y_1'}{y_1} \right].$$

Now, for any point (x, y) in the image we have, using (7) with $\hat{\beta} = 0$:

$$\frac{x'}{x} - \frac{y'}{y} = \frac{\Delta x}{x} - \frac{\Delta y}{y} = \hat{\alpha} \cdot I \cos(\nu)/R = \hat{\alpha} \cdot \tan(2\mu).$$

This leads to the final equations:

$$\tan(2\mu) = \frac{1}{x_2} \cdot \left[\frac{y_2'}{y_2} - \frac{y_1'}{y_1} \right]. \tag{9}$$

and:

$$\hat{\alpha} = \frac{\frac{x'}{x} - \frac{y'}{y}}{\tan(2\mu)} \; ; \quad \tan(\nu) = \frac{\frac{\Delta y_1}{y_1}}{\tan(2\mu)} \; ; \quad R = \frac{I}{\sqrt{\tan^2(2\mu) + \frac{\Delta y_1}{y_1}^2}}.$$

The ratio $\frac{y'}{y}$ near the vertical meridian is relatively reliable and easy to obtain. However, a point with $\hat{\alpha} \approx 0$ does not necessarily exist, in which case we can:

1. Follow Mayhew & Longuet-Higgins (1982) and neglect the term $\hat{\alpha}\tan(\nu)$, but not the term $\tan(\nu)$,

$$\tan(2\mu) = \frac{1}{x} \cdot \left[\frac{y'}{y} - \frac{y_1'}{y_1} \right]. \tag{9'}$$

If we neglect $\tan(\nu)$, for consistency, we obtain:

$$\tan(2\mu) = \frac{1}{x} \cdot \frac{\Delta y'}{y}. \tag{9''}$$

10

2. Solve the initial scheme without such a point. Given a vertical-meridian point, there remains a fairly simple equation to solve. This would be:

$$\tan(2\mu) - \hat{\alpha} \cdot \frac{\Delta y_1{}'}{y_1} = \frac{1}{x} \cdot \left(\frac{y'}{y} - \frac{y_1'}{y_1}\right)$$

$$\tan(2\mu) \cdot \hat{\alpha} = \frac{x'}{x} - \frac{y'}{y} , \tag{10}$$

which reduces, after substituting $\tan(2\mu)$ from the second equation in the first equation, to a second degree polynomial in $\hat{\alpha}$.

A different numerical approach would be to use, for example, three images taken while moving on the base plane. Denote by (x_0, y_0), (x_1, y_1) and (x_2, y_2) the coordinates of the conjugate projections of some point P on the three images. Denote by α_1, μ_1 and ν_1 the angle of tilt, half the angle of convergence, and the angle of gaze, respectively, in the coordinate system defined as above by the first two images. Denote by α_2, μ_2 and ν_2 the same angles in the coordinate system defined by the last two images (see figure 3). For motion on a straight line we have:

$$\tan\alpha_1 \cdot \tan\mu_1 = \frac{\lambda(x_0, y_0, x_1, y_1) - 1}{\lambda(x_0, y_0, x_1, y_1) + 1}$$

$$\tan\nu_1 \cdot \tan\mu_1 = \frac{1 - \frac{y_1(x=0)}{y_0(x=0)}}{1 + \frac{y_1(x=0)}{y_0(x=0)}}$$

$$\tan\alpha_2 \cdot \tan\mu_2 = \frac{\lambda(x_1, y_1, x_2, y_2) - 1}{\lambda(x_1, y_1, x_2, y_2) + 1}$$

$$\tan\nu_2 \cdot \tan\mu_2 = \frac{1 - \frac{y_2(x=0)}{y_1(x=0)}}{1 + \frac{y_2(x=0)}{y_1(x=0)}}$$

$$\alpha_1 - \alpha_2 = \mu_1 + \mu_2$$

$$\nu_2 - \nu_1 = \mu_1 + \mu_2,$$

where $\frac{y_j(x=0)}{y_i(x=0)}$ is the Y-axis ratio on the vertical meridian ($x = 0$) between conjugate points in images i and j.

Thus we have six nonlinear equations with six unknowns. For small μ_i's we have approximately a linear problem, where the solution is a null vector of the approximating matrix. We will obtain a similar set of equations if we take two points in the three images and ignore the equations involving the angle of gaze ν. In this case the motion in the base plane does not have to be in a straight line.

Error Analysis

First, from the definition of λ and χ it follows that the base plane itself is singular in
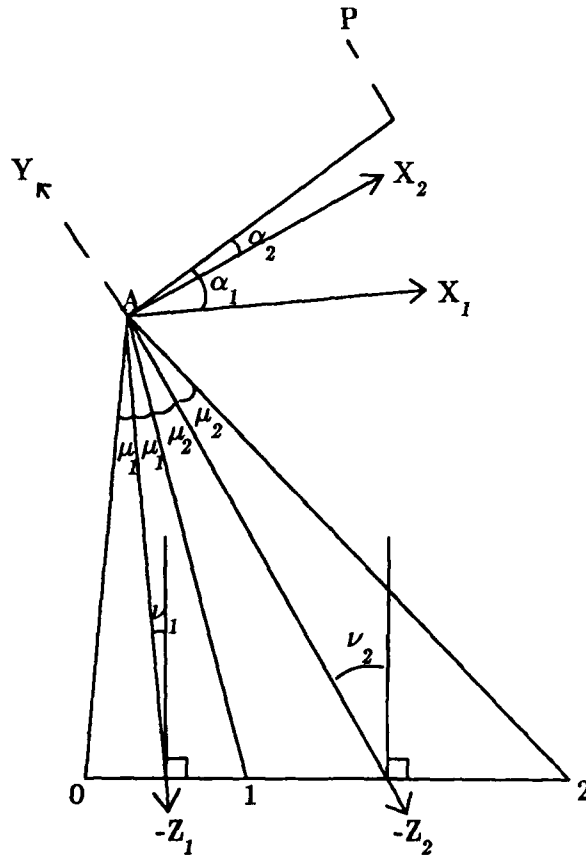
Figure 3. Three images from motion on the base plane.

the sense that these orders are not defined for points on it. The same problem exists in the analysis of normals to surfaces of objects. One can, however, estimate the orders and normals by substituting $\frac{y_r}{y_l}$ of a matched point far from the base plane. More specifically, for $P = (x, y, z)$ we have

$$
\frac{y_r}{y_l} = \frac{d_l}{d_r} + \frac{z}{d_r} \frac{2 \sin \mu \tan \nu}{1 + \tan \mu \tan \nu} + \frac{x}{d_r} \frac{2 \sin \mu}{1 + \tan \mu \tan \nu} + o(\frac{x}{d_r}, \frac{z}{d_r})
$$

$$
= 1 - \frac{2 \tan \mu \tan \nu}{1 + \tan \mu \tan \nu} + \frac{z}{d_r} \frac{2 \sin \mu \tan \nu}{1 + \tan \mu \tan \nu} + \frac{x}{d_r} \frac{2 \sin \mu}{1 + \tan \mu \tan \nu} + o(\frac{x}{d_r}, \frac{z}{d_r}).
$$

Thus, if point P^j is used to approximate point P^i, the error will be:

$$
\left(\frac{y_r}{y_l} \right)^i - \left(\frac{y_r}{y_l} \right)^j = \frac{2 \sin \mu}{1 + \tan \mu \tan \nu} \left[\frac{z^i - z^j}{d_r} \tan \nu + \frac{x^i - x^j}{d_r} \right].
$$

The error is 0 if the approximating point P^j lies exactly "above" P^i (differs only in the

y-coordinate).

Moreover, one can use as an estimation $\frac{1-\tan\mu\,\tan\nu}{1+\tan\mu\,\tan\nu}$ (the first two terms), so that some (possibly extraretinal) estimation of μ (half the angle of convergence) and ν (the angle of gaze) will suffice to give a rough estimation to $\frac{y_r}{y_l}$ when no other source of information is available. Note that one can not take $\frac{y_r}{y_l} \approx 1$ when computing $x_r - \frac{y_r}{y_l}x_l$, as a first order approximation in μ, since $x_r - x_l$ is of the order of magnitude of μ also.

Second, let us consider the violation of the basic assumption, namely, that the X-axes themselves of both cameras are epipolar lines. This introduces an error δ_r and δ_l in the polar angle of a given point's projections on the right and left images, respectively. Thus, the true orders should be:

$$\lambda = \frac{\cot(\vartheta_r + \delta_r)}{\cot(\vartheta_l + \delta_l)}$$

$$= \frac{\cot\vartheta_r}{\cot\vartheta_l} + \frac{\cot\vartheta_r}{\cos^2\vartheta_l}\cdot\delta_l - \frac{\tan\vartheta_l}{\sin^2\vartheta_r}\cdot\delta_r + o(\delta),$$

$$\chi^y = \cot(\vartheta_r + \delta_r) - \cot(\vartheta_l + \delta_l)$$

$$= (\cot\vartheta_r - \cot\vartheta_l) - \frac{1}{\sin^2\vartheta_r}\cdot\delta_r + \frac{1}{\sin^2\vartheta_l}\cdot\delta_l + o(\delta),$$

where $\delta_r, \delta_l \leq \delta$. The main conclusion from this is that the effect of axes misalignment is greater near the horizontal and vertical meridians, and possibly negligible further away. Also, this error affects less the expressions for qualitative shape (the normal to iso-x or iso-y surface contours), since they involve differences where this error is somewhat cancelled out for the two points.

Discussion: Comparison to some empirical data

The orders λ and χ as defined above, and the way the scaling coefficients depend on camera parameters, seem to be consistent with the following psychophysical results:

1. The advantage of relative depth perception in human vision, which is more reliable than absolute depth perception. That is, it is easier to distinguish between different objects if they differ a little in depth than to give a good estimation to the absolute depth of a given object in space with no additional information of perspective.

2. The induced effect, as discussed above, which can be shown to be a natural side effect of using the tilt-related order λ to estimate the tilt of a plane at the center of gaze. No assumptions on the way the visual system finds and interprets corresponding points is needed. Moreover, this is a "local" explanation of the induced effect in the sense that it allows for opposite induced effects in neighboring spatial regions, in agreement with psychophysical evidence (see Rogers & Koenderink, 1986). Likewise, this explanation does not imply a perceived asymmetric convergence of the eyes, again in agreement with empirical data. It is interesting to note that λ might be the discrete equivalent to

13

the term of the optical flow field used by Rogers & Koenderink (1986) to explain the induced effect with motion parallax. Quantitatively, λ is more susceptible to errors near the meridians, in agreement with psychophysical experiments (Arditi, 1982) that show deterioration in human performance of tilt estimation near the meridians. Finally, such an effect for a plane not passing through the fixation point is predicted by our qualitative shape analysis.

3. The difficulty in comparing right-hemifield and left-hemifield points for their depth, as demonstrated by the following experiment (Rogers & Graham, 1983): given two planes, one tilted away from the viewer from the center to the right and one tilted towards the viewer from the center to the left, whose end points on the left and right sides respectively are equidistant, the observer will (wrongly) perceive the right side of the first plane as closer in depth than the left side of the second plane. If, however, the whole configuration is rotated by 90° so that one plane is tilted upwards from the center and the other is tilted downwards from the center, with far from the center end sides equidistant, then the observer will actually (rightly) perceive those two ends as equidistant. This result is consistent with the above order λ, which orders points on the right and left hemifields separately, on either side of the horizontal meridian, but not if they are on different sides of the vertical meridian. χ is also defined differently in the two visual hemifields, and it has some distortion as a function of the horizontal distance between the two points.

4. Empirical evidence for the dependence of relative depth perception on extraretinal perception of the angle of convergence of the eyes, which is predicted by using the orders λ and χ to evaluate depth with no additional computation.

5. The anisotropy between the horizontal and vertical dimensions in relative depth perception, as demonstrated, e.g., by the advantage of using flanking lines as test objects versus vertically displaced lines (Westheimer, 1979). Moreover, both vertically displaced or horizontally displaced stimuli are preferred over diagonally displaced stimuli in the sense of lower disparity threshold for discrimination. This is consistent with two distinct qualitative orders that prefer horizontal or vertical displacement.

6. Psychophysical results suggesting either deterioration in depth distinction when points are coplanar with the point of fixation or depth perception relative to a plane through the fixation point, which is context dependent and not necessarily the fronto-parallel plane (see Mitchison & Westheimer, 1984). Note that λ is constant for points coplanar with the fixation point.

7. The use of λ for threshold estimation in acuity experiments predicts some results like an increase of threshold with standing disparity (Westheimer, 1979) or the cosine rule (Ebenholz & Walchli, 1965).

Moreover, λ and χ^y depend only on the polar angles of the conjugate points in both images, a quantity that is preserved under projection to a spherical body (an eye) or a planar body (a camera). It is interesting to note, in this respect, that the first visual transformation from the retina to $V1$ in primates seems to be in good agreement with

14

the complex-log mapping (Schwartz, 1984), namely: $(x, y) \longrightarrow (\log r, \vartheta)$. This mapping explicitly computes the polar angle ϑ of a point.

Summary

The goal of this work had been to obtain qualitative information from a stereo pair, with as few computations as possible and minimal dependence on the camera and scene parameters. We have shown that points in a stereo pair, once matched to each other, can be ordered according to two distinct order expressions: a tilt-related order λ, which is roughly a relative depth order when ordering points with only vertical displacement, and a depth-related order χ which is best to order points with only horizontal displacement. These orders are completely determined by image coordinates of conjugate points, and no camera or scene parameters are needed (which need not be similar for both cameras). λ and some variation of χ (χ^y) depend only on the polar angles of the conjugate points in both images, a quantity which is preserved under projection to a spherical body (an eye) or a planar body (a camera). Moreover, given the polar angles of the images coordinates, some qualitative shape information can be obtained: one can follow changes in the curvature of a contour on the surface of any object in the field of view, with either x- or y-coordinate fixed. We demonstrated, by further analyzing the exact equations, that obtaining the quantitative information is much harder and less reliable than obtaining the qualitative one, e.g., orders like λ and χ. It usually involves some assumptions on the scene or extra-retinal information, plus a lot of computations. These computations tend to be less robust and sensitive to noise and errors. Finally, we discussed some psychophysical (and neurophysiological) data which seem to support the use of such orders in humans. Most notably, the use of λ for tilt estimation predicts "the induced size effect", an unusual behavior of the human visual system which lacks other straightforward explanation.

Appendix

Consider the base plane, which includes the X-axes of both cameras and both their optical-axes. This is illustrated in figure 4, where O is the focal node of one camera, A is the fixation point, A' is the projection of A on the image plane or the origin of the camera coordinate system, B is the projection of a given point in space (C) on the base plane, B' is the projection of B on the camera X-axis, and C' is the projection of the point C on the image plane. In the base plane, we add the point D which is the projection of B on the optical axis $\overline{AA'}$. Let φ denote the angle $\angle BAO$. Let (x, y) denote the coordinates of the projection of point C on the image, so that $x = \overline{A'B'}$ and $y = \overline{B'C'}$. Let d denote the distance of the fixation point to the eye, so that $d = \overline{AO}$, and let h denote the focal length

of the camera, so that $h = \overline{A'O}$. Using similar triangles, one can verify the following:

$$\frac{\overline{BC}}{y} = \frac{\overline{BO}}{\overline{B'O}} = \frac{\overline{DO}}{\overline{A'O}} = \frac{d - \overline{AB}\cos\varphi}{h}.$$
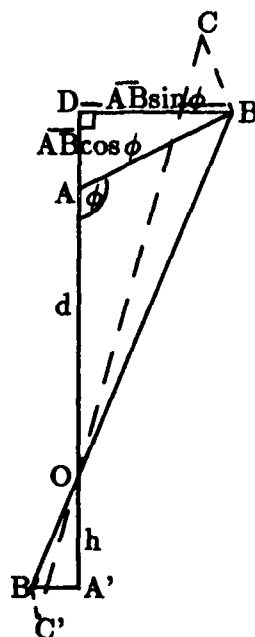
Figure 4. The base plane viewed from above, with one camera.

In other words,

$$y = \frac{h\overline{BC}}{d - \overline{AB}\cos\varphi}. \tag{11}$$

Using the same arguments, we obtain

$$\frac{\overline{AB}\sin\varphi}{x} = \frac{\overline{DO}}{\overline{A'O}} = \frac{d - \overline{AB}\cos\varphi}{h}.$$

In other words,

$$x = \frac{h\overline{AB}\sin\varphi}{d - \overline{AB}\cos\varphi}.$$

Thus

$$\frac{x}{y} = \frac{\overline{AB}}{\overline{BC}} \cdot \sin\varphi.$$

Note that our assumption that the base plane intersects both cameras' X-axes implies that the same geometry holds for both cameras in the sense that the segments $\overline{BC}$ and $\overline{AB}$ are identical in both cases. (A and C are the same points, and B is identical since C is projected

16

onto the same plane.) Let us add indices for the variables of the left and right cameras, l and r respectively. Then

$$\frac{x_r}{y_r} = \frac{\overline{AB}}{\overline{BC}} \cdot \sin\varphi_r$$

$$\frac{x_l}{y_l} = \frac{\overline{AB}}{\overline{BC}} \cdot \sin\varphi_l. \tag{12}$$

Finally

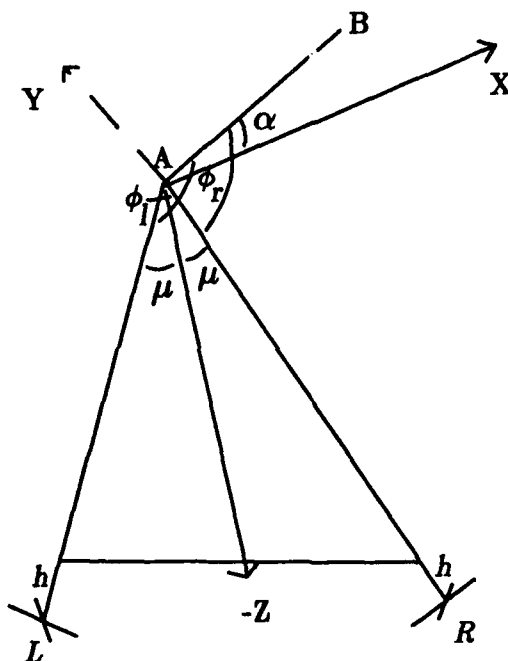$$\frac{x_r}{y_r} / \frac{x_l}{y_l} = \frac{\sin\varphi_r}{\sin\varphi_l}. \tag{13}$$

Figure 5. Angles of tilt on the base plane as viewed from above.

Figure 5 illustrates the geometry on the base plane with both cameras, and the coordinate system used in the text. Recall that 2μ denotes the angle of convergence of the cameras in the base plane, and that the Z-axis is defined as the angle-bisector of this angle. Also, α was defined as $\arctan(\frac{z}{x})$, so that $\alpha = \frac{\varphi_r + \varphi_l}{2} - 90^\circ$. These definitions of μ and α imply the following:

$$\varphi_r = \alpha + 90^\circ - \mu$$

$$\varphi_l = \alpha + 90^\circ + \mu.$$

Thus

$$\frac{\sin\varphi_r}{\sin\varphi_l} = \frac{\cos\alpha\cos\mu + \sin\alpha\sin\mu}{\cos\alpha\cos\mu - \sin\alpha\sin\mu} = \frac{1 + \tan\alpha\tan\mu}{1 - \tan\alpha\tan\mu}. \tag{14}$$

We have defined $\lambda = \frac{\cot\vartheta_r}{\cot\vartheta_l} = \frac{x_r}{y_r}/\frac{x_l}{y_l}$. From (13) and (14) we obtain

$$\lambda = \frac{1 + \tan\alpha\tan\mu}{1 - \tan\alpha\tan\mu} \implies \tan\alpha\tan\mu = \frac{\lambda - 1}{\lambda + 1},$$

which is equivalent to equation (1).

Now, if $C = (x, y, z)$ in the world coordinate system we have defined above, then

$$\cot\vartheta_r - \cot\vartheta_l = \frac{x_r}{y_r} - \frac{x_l}{y_l} = \frac{\overline{AB}}{\overline{BC}}(\sin\varphi_r - \sin\varphi_l)$$

$$= \frac{\overline{AB}}{\overline{BC}}2\sin\alpha\sin\mu$$

$$= 2\sin\mu\frac{z}{y}.$$

Since by definition $\tan\beta = \frac{z}{y}$, we immediately obtain equation (2).

Let us develop the expression for the image coordinate y above, considering the right image with no loss of generality. We then have from (11),

$$y_r = \frac{h\overline{BC}}{d_r - \overline{AB}\cos\varphi_r}$$

$$= \frac{hy}{d_r - x\sin\mu + z\cos\mu}.$$

From figure 1, in which the angle of gaze ν is defined, it follows that $d_r = \frac{I\cos(\mu - \nu)}{\sin 2\mu}$, so that

$$y = \left(\frac{I\cos(\mu - \nu)}{\sin 2\mu} - x\sin\mu + z\cos\mu\right)\cdot\frac{y_r}{h}. \tag{6}$$

Applying the same argument to the y coordinate of the left image, we obtain

$$y = \left(\frac{I\cos(\mu + \nu)}{\sin 2\mu} + x\sin\mu + z\cos\mu\right)\cdot\frac{y_l}{h}. \tag{6'}$$

References

Arditi, A., The dependence of the induced effect on orientation and a hypothesis concerning disparity computations in general, *Vision Res.* V-22, 247-256, 1982.

Arditi, A., L. Kaufman, and J. A. Movshon, A simple explanation of the induced size effect, *Vis. Res.* V-21, 755-764, 1981.

Ebenholz, S., and R. Walchli, Stereoscopic thresholds as a function of head- and object-orientation, *Vision Res.* V-5, 455-461, 1965.

Foley, J. M., Primary distance perception, in: *Handbook of sensory physiology*, 1977.

Foley, J.M., and W. Richards, Effects of voluntary eye movement and convergence on the binocular appreciation of depth, *Perception & psychophysics* V-11(6), 423-427, 1972.

Horn B. K. P., "Robot Vision", MIT Press & McGraw-Hill, 1986.

Longuet-Higgins H. C., A computer algorithm for reconstructing a scene from two projections, *Nature* 293, 133-135, 1981.

Longuet-Higgins H. C., The role of vertical dimension in stereoscopic vision, *Perception* V-11, 377-386, 1982.

Longuet-Higgins H. C., and K. Prazdny, The interpretation of a moving retinal image, *Proc. R. Soc. Lond. B* 208, 385-397, 1980.

Mayhew, J. E. W., The interpretation of stereo-disparity information: the computation of surface orientation and depth, *Perception* V-11, 387-403, 1982.

Mayhew, J. E. W., and H. C. Longuet-Higgins, A computational model of binocular depth perception, *Nature* V-297 (3), 376-378, 1982.

Mitchison, G. J., and G. Westheimer, The perception of depth in simple figures, *Vision Res.* V-24, No-9, 1063-1073, 1984.

Ogle, K. N., Induced size effect. I. A new phenomenon in binocular space perception associated with the relative sizes of the images of the two eyes, *Archs Ophthal* 20, 604-623, 1938.

Prazdny, K., Stereoscopic matching, eye position, and absolute depth, *Flair Draft*, 1981.

Rogers, B. J., and M. E. Graham, Anisotropies in the perception of three-dimensional surfaces, *Science* V-221, 1409-1411, 1983.

Rogers, B. J., and J. Koenderink, Monocular aniseikonia: a motion parallax analogue of the disparity-induced effect, *Nature* V-322(3), 62-63, 1986.

Schwartz, E. L., Anatomical and physiological correlates of visual computation from striate to infero-temporal cortex, *IEEE Trans. on Systems, Man, and Cybernetics* SMC-14 (2), 257-271, 1984.

Westheimer, G., Cooperative neural processes involved in stereoscopic acuity, *Exp. Brain Res.* 36, 585-597, 1979.